Deep Unfolding-Based Channel Estimation for IRS-Aided mmWave Systems via Two-Stage LAMP Network with Row Compression

Wen-Chiao Tsai, *Member, IEEE*, Chi-Wei Chen, *Student Member, IEEE*, and An-Yeu (Andy) Wu, *Fellow, IEEE*

***Abstract*—Intelligent reflecting surface (IRS) is emerging as a promising and revolutionary technology for achieving cost-effective wireless communication systems. However, the performance of IRS communications heavily relies on acquiring accurate channel state information (CSI), which is challenging under low training overhead due to a large number of passive IRS elements. This paper proposes a two-stage LAMP network with row compression (RCTS-LAMP) to solve the joint estimation problem of direct and cascaded channels in IRS-aided millimeter-wave (mmWave) systems. Specifically, the proposed RCTS-LAMP is a model-driven neural network that combines the advantages of compressive sensing (CS) and deep learning (DL) by using the deep unfolding technique. By doing so, we can recover the direct and cascaded channels with CS under low training overhead, and the estimation performance can be significantly improved with the joint optimization of DL. Meanwhile, the cascaded channel estimation is decomposed into two stages to reduce computational complexity further. Numerical results show that the RCTS-LAMP network can estimate the cascaded channel with a better trade-off between computational complexity and accuracy, while the direct channel can be jointly recovered without adding an extra network.**

***Index Terms*—Intelligent reflecting surface (IRS), channel estimation, compressive sensing, deep unfolding.**

I. Introduction

RECENTLY, intelligent reflecting surface (IRS), also known as reconfigurable intelligent surface (RIS), is emerging as a promising technology for improving wireless communication performance with a variety of applications, such as coverage extension, improving channel rank condition, and interference suppression [1]–[4]. In IRS communications, a surface consisting of a large number of low-cost reflective elements is used to control the propagation of radio waves by adjusting the amplitudes and phase shifts of the reflected signals. This technology offers the potential to create a virtual line-of-sight (LoS) link to bypass obstacles between transceivers and provide additional beamforming gain to achieve robust millimeter-wave (mmWave) communications [5], [6]. Besides, a large IRS hardens the end-to-end channel of a multiple-input multiple-output (MIMO) system with a few transmit and receive antennas [7], which reduces the need for large-scale antenna arrays. However, the performance of IRS communications heavily depends on accurate channel state information (CSI) for optimizing system performance.

That is, accurate channel estimation is crucial for effective communications in IRS systems. It enables the IRS to adjust the phase shifts of its reflective elements to maximize the system objective. For the uplink channel training, the channel estimation process typically involves transmitting known pilots from the user, which are reflected by the IRS and received by the base station (BS). However, channel estimation in fully passive IRS-aided systems is challenging since the user-IRS and IRS-BS channels cannot be estimated separately by the passive IRS elements without sensing capabilities. Instead, only the cascaded channel from the user to the BS through the IRS can be estimated. In addition, cascaded channel estimation requires large training overhead due to the large number of IRS elements, which makes traditional channel estimation methods computationally expensive and time-consuming. For instance, least square (LS)-based methods [8], [9] are proposed to estimate the cascaded channel. Although the optimal training reflection pattern at the IRS and the pilot sequence at the user are derived in [8] and [9], these methods require the number of training subframes to be larger or equal to the number of IRS elements, which may hinder the potential performance gain of IRS. Hence, several techniques have been proposed to address these challenges, such as *compressive sensing (CS)* and *deep learning (DL)*-based methods.

For IRS-aided mmWave systems, the mmWave channels are composed of a limited number of paths due to severe path loss over distance. Thus, the mmWave channels usually exhibit row-rank nature and strong sparsity in the angular domain, which can be exploited to reduce the training overhead. Specifically, by formulating the cascaded channel estimation problem as a sparse signal recovery problem [10], [11], several CS-based algorithms have been developed for IRS channel estimation, such as orthogonal matching pursuit (OMP)-based algorithms [10]–[12], and approximate message passing (AMP)-based algorithms [13], [14]. Besides, by exploiting the common sparsity in IRS-aided multi-user systems, [15] proposes a two-step multi-user joint channel estimation procedure for efficient recovery. Although these methods are able to recover the cascaded channel with reduced training overhead, better estimation performance comes at the expense of higher grid resolutions in the virtual angular domain (VAD). Since both the estimation performance and computational complexity increase with the grid resolutions, the cost-performance trade-off is crucial for CS-based algorithms. Although the grid mismatch problem can be solved with off-grid CS methods, such as atomic norm minimization [16], these methods are

This work is supported in part by MediaTek, Inc., Hsinchu, Taiwan, under Grant MTKC-2023-1050, and in part by the Ministry of Science and Technology of Taiwan under Grant MOST 110-2221-E-002-184-MY3 (Corresponding author: An-Yeu (Andy) Wu).

The authors are with the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, 10617, Taiwan (e-mail: daniel@access.ee.ntu.edu.tw; wilbur@access.ee.ntu.edu.tw; andywu@ntu.edu.tw).

often computationally demanding and require the number of paths as prior information, which is not practical for real-time applications.

On the other hand, deep learning (DL)-based methods [17]–[19] have shown significant potential in channel estimation for IRS-aided systems. The authors in [17] employ the deep denoising neural network (NN) to improve the performance of the CS estimator in the hybrid passive/active IRS architecture. By exploiting the low-rank nature of the mmWave channel, the training overhead is reduced while the performance is enhanced with the denoising network. However, the additional active channel sensors increase the hardware complexity. In [18] and [19], the cascaded channel estimation is modeled as a denoising problem. The proposed CNN-based networks adopt the noisy LS-based solution as input to perform the denoising function. Nonetheless, these works only focus on improving estimation accuracy, and the critical issue of large training overhead is not addressed. Therefore, DL-based methods for training overhead reduction and achieving reliable estimation accuracy are essential for passive IRS-aided systems.

To estimate the cascaded channel with reduced training overhead, [20] and [21] propose to activate only a portion of IRS elements and predict the full channel from the partial one. Specifically, the synthetic deep neural networks (DNNs) in [20] and [21] sequentially perform the estimation of direct and active cascaded channels and the prediction for the full cascaded channel. However, training generic DNNs for channel estimation and prediction may suffer from the issues of interpretability and generalizability. In contrast to the generic data-driven NNs, a promising technique called deep unfolding has been developed for model-driven NNs that fuses conventional algorithms of certain performance guarantees with DL tools. Specifically, deep unfolding introduces tunable parameters into an iterative algorithm, where the iterations are unfolded into the layers of an NN, and the tunable parameters can be optimized with DL. Due to better interpretability and generalizability, deep unfolding has shown its effectiveness in various communication systems, such as MIMO detection [22] and channel decoding [23], [24].

In this paper, we extend our previous work [25] to study the joint estimation problem of direct and cascaded channels for passive IRS-aided mmWave systems. To achieve a better trade-off between estimation accuracy and computational complexity, we utilize the promising deep unfolding technique that combines the advantages of CS and DL. Specifically, to enhance the performance of CS-based methods under low grid resolutions, we exploit the learned AMP (LAMP) network [26], which unfolds the iterative AMP algorithm. By optimizing the linear transform matrices and shrinkage parameters of the LAMP network, we can improve the channel estimation performance. However, two issues should be addressed:

1) *High computational complexity for joint CS recovery:* Due to the improved estimation performance, the LAMP network can avoid recovering the channel under high grid resolutions. However, the computational complexity may be prohibitive due to the large-size 3D measurement matrix in the joint CS problem that directly recovers the angular cascaded channel. Thus, joint CS recovery should be avoided due to the computational complexity.
2) *Separate networks for the estimation of direct and cascaded channels:* Since both the estimations of the direct and cascaded channels can be formulated as sparse signal recovery problems, we can solve each problem with a LAMP network to achieve high estimation accuracy. However, separate LAMP networks result in additional training and memory overhead, which may hinder the potential for practical deployment.

By utilizing the inherent structure of the angular cascaded channel, we propose a two-stage LAMP network with row compression (RCTS-LAMP), which can significantly reduce the complexity of cascaded channel estimation and avoid an extra network for direct channel estimation. Our main contributions are summarized as follows:

1) *Design of two-stage LAMP (TS-LAMP) network:* We decompose the cascaded channel estimation process into two stages for complexity reduction. To solve the two sparse signal recovery problems in the two stages, we use two LAMP networks for sequential recovery of a row-sparse matrix and the angular cascaded channel. By doing so, the consequent TS-LAMP network can significantly reduce the execution time by 83% and improve the estimation accuracy compared with the joint CS recovery using one LAMP network.
2) *Row compression mechanism based on row sparsity:* By exploiting the row-structured sparsity of the angular cascaded channel, we propose a row compression mechanism to compress the row-sparse matrix and reduce the complexity in the second stage. Specifically, since each row of the row-sparse matrix corresponds to an angle of arrival (AoA) at the BS, we can prune the low-power rows based on the angular-domain sparsity at the BS. Simulation results show that the proposed RCTS-LAMP network can maintain the performance of TS-LAMP with a further reduction of execution time from 40% to 47%.
3) *Learned vector shrinkage function for the multiple measurement vector (MMV) problem:* Note that the recovery of the row-sparse matrix is an MMV problem whose columns exist common sparsity. Hence, we propose a LAMP-MMV network to recover the row-sparse matrix from the MMV perspective, where a learned vector shrinkage function is employed to utilize the common sparsity. As a result, the RCTS-LAMP-MMV network integrates the LAMP-MMV in the first stage and achieves better estimation performance without much complexity increase compared with the RCTS-LAMP network.
4) *Three-stage training procedure using transfer learning:* Since the RCTS-LAMP network can recover the angular domain signal at the BS in the first stage, the knowledge learned from recovering the row-sparse matrix can be utilized to recover the sparse angular direct channel. Therefore, we propose a three-stage training procedure for the joint estimation of the direct and cascaded channels, where the direct channel can be recovered with the first LAMP network in RCTS-LAMP through transfer

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

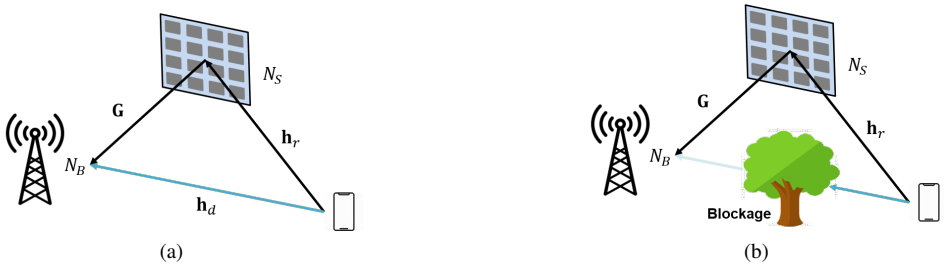


Fig. 1. IRS-aided single-user wireless communication systems: (a) the system with direct channel, (b) the system without direct channel due to blockage.

learning. As a result, a separate network with extra training and memory overhead can be avoided.

The rest of this paper is organized as follows. Section II introduces the system model and formulation of the direct and cascaded channel estimation problems. Section III provides the details about the proposed RCTS-LAMP network, followed by the three-stage training procedure for the joint estimation of direct and cascaded channels in Section IV. The numerical experiments and complete analyses are shown in Section V. Finally, Section VI concludes our work.

Notations: Scalars are denoted by italic letters, vectors and matrices are denoted by lower and upper-case boldface letters, respectively. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\|\mathbf{A}\|$, and vec$(\mathbf{A})$ are the transpose, conjugate transpose, inverse, Frobenius norm, and vectorization of matrix $\mathbf{A}$, respectively. $\|\mathbf{x}\|$ denotes the l_2-norm of vector $\mathbf{x}$, while diag$(\mathbf{x})$ denotes the diagonal matrix with $\mathbf{x}$ as its main diagonal elements. The imaginary unit is denoted by $j \equiv \sqrt{-1}$. The Kronecker product and the Khatri–Rao product are represented by $\otimes$ and $\odot$, respectively. Finally, $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ denotes the circularly symmetric complex Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma}$, and $\mathbb{E}\{\cdot\}$ denotes the expectation operator.

II. System Model and Problem Formulation

In this section, we will introduce the system and channel models of the IRS-aided mmWave communication systems. Then, we will formulate the channel estimation problems for the direct and cascaded channels.

A. System Model and Channel Model

Consider the IRS-aided mmWave communication systems shown in Fig. 1. We assume that the BS equipped with an N_B-antenna uniform linear array (ULA) serves a single-antenna user with the assistance of an IRS with an N_S-element uniform planar array (UPA). As in most mmWave systems, the BS employs a hybrid analog-digital precoding architecture with $N_{RF}^B \leq N_B$ RF chains to reduce hardware costs and power consumption. By smartly reconfiguring the IRS reflection coefficients, IRS can enhance the communication systems under various channel conditions. For instance, IRS can improve the channel rank condition by establishing an additional virtual LoS link between transceivers, as demonstrated in Fig. 1(a). Besides, reliable communications can be achieved by utilizing the virtual LoS link to bypass the obstacles between transceivers, as depicted in Fig. 1(b).

We assume the widely used Saleh-Valenzuela (SV) channel model [27], which is based on the angle of departures (AoDs), the AoAs, and the complex path gains of each link. Let $\mathbf{G} \in \mathbb{C}^{N_B \times N_S}$ denote the channel from the IRS to the BS, $\mathbf{h}_r \in \mathbb{C}^{N_S \times 1}$ denote the channel from the user to the IRS, and $\mathbf{h}_d \in \mathbb{C}^{N_B \times 1}$ denote the direct channel from the user to the BS. Suppose $\mathbf{G}$, $\mathbf{h}_r$ and $\mathbf{h}_d$ are characterized by L_G, L_r and L_d propagation paths, respectively. According to the SV channel model $\mathbf{G}$, $\mathbf{h}_r$ and $\mathbf{h}_d$ can be modeled as

$$\mathbf{G} = \sqrt{\frac{N_B N_S}{L_G}} \sum_{l_G=1}^{L_G} \alpha_G^{l_G} \mathbf{a}_B\left(\varphi_G^{l_G}\right) \mathbf{a}_S^H\left(\vartheta_G^{l_G}, \gamma_G^{l_G}\right), \tag{1}$$

$$\mathbf{h}_r = \sqrt{\frac{N_S}{L_r}} \sum_{l_r=1}^{L_r} \alpha_r^{l_r} \mathbf{a}_S\left(\vartheta_r^{l_r}, \gamma_r^{l_r}\right), \tag{2}$$

$$\mathbf{h}_d = \sqrt{\frac{N_B}{L_d}} \sum_{l_d=1}^{L_d} \alpha_d^{l_d} \mathbf{a}_B\left(\varphi_d^{l_d}\right), \tag{3}$$

where $\alpha_G^{l_G}$, $\alpha_r^{l_r}$ and $\alpha_d^{l_d}$ are the complex path gains of the corresponding paths, $\varphi_G^{l_G}$ and $\varphi_d^{l_d}$ denote the AoA spatial frequencies at the BS, $\vartheta_G^{l_G}$ and $\gamma_G^{l_G}$ ($\vartheta_r^{l_r}$ and $\gamma_r^{l_r}$) represent the azimuth and elevation AoD (AoA) spatial frequencies at the IRS. The spatial frequencies are defined as $\varphi = 2\pi d \sin(\varphi_a)/\lambda$, $\vartheta = 2\pi d \sin(\vartheta_a)\cos(\gamma_a)/\lambda$, and $\gamma = 2\pi d \sin(\gamma_a)/\lambda$, where d denotes the antenna spacing, λ is the signal wavelength, $\varphi_a \in [-\pi/2, \pi/2]$ denotes the path angle of $\varphi \in \left\{\varphi_G^{l_G}, \varphi_d^{l_d}\right\}$ at the BS in the angular domain, $\vartheta_a \in [-\pi/2, \pi/2]$ ($\gamma_a \in [-\pi/2, \pi/2]$) denotes the azimuth (elevation) path angle of $\vartheta \in \left\{\vartheta_G^{l_G}, \vartheta_r^{l_r}\right\}$ ($\gamma \in \left\{\gamma_G^{l_G}, \gamma_r^{l_r}\right\}$) at the IRS in the angular domain. Moreover, $\mathbf{a}_B(\varphi) \in \mathbb{C}^{N_B \times 1}$ and $\mathbf{a}_S(\vartheta, \gamma) \in \mathbb{C}^{N_S \times 1}$ are the array response vectors at the BS and the IRS, respectively. For the N_B-antenna ULA, $\mathbf{a}_B(\varphi)$ can be represented by

$$\left[\mathbf{a}_B(\varphi)\right]^T = \sqrt{\frac{1}{N_B}}\left[1, e^{j\varphi}, \ldots, e^{j(N_B-1)\varphi}\right], \tag{4}$$

For the N_S-element UPA with N_S^h horizontal elements and N_S^v vertical elements, $\mathbf{a}_S(\vartheta, \gamma)$ can be represented by

$$\begin{aligned}\left[\mathbf{a}_S(\vartheta, \gamma)\right]^T = &\sqrt{\frac{1}{N_S}}\left[1, \ldots, e^{j\left(N_S^h-1\right)\vartheta}\right] \otimes \\ &\left[1, \ldots, e^{j(N_S^v-1)\gamma}\right].\end{aligned} \tag{5}$$

Furthermore, we denote $\mathbf{H}_c = \mathbf{h}_r^T \odot \mathbf{G} \in \mathbb{C}^{N_B \times N_S}$ as the cascaded channel. Then, the effective channel $\mathbf{H} \in \mathbb{C}^{N_B \times 1}$

for the user-IRS-BS link considering the reflection of IRS becomes

$$\mathbf{H}\left(\boldsymbol{\theta}\right)=\mathbf{G}\mathrm{diag}(\boldsymbol{\theta})\mathbf{h}_r=\mathbf{H}_c\boldsymbol{\theta}, \tag{6}$$

where $\boldsymbol{\theta}=\left[s_1e^{j\phi_1},\ldots,s_{N_S}e^{j\phi_{N_S}}\right]^T$ is the reflection vector at the IRS, with $s_n\in\{0,1\}$ and $\phi_n\in[0,2\pi)$ representing the on/off state and the phase shift of the n^{th} IRS element. In practice, each IRS element can be adjusted through a PIN diode, and the reflective element is almost transparent, with the insertion loss nearly zero when the PIN diode is turned off [28]. Based on (1) and (2), $\mathbf{H}_c$ can also be represented by the SV channel model, which is derived as

$$\begin{aligned}\mathbf{H}_c=\sqrt{\frac{N_BN_S}{L_GL_r}}\sum_{l_G=1}^{L_G}\sum_{l_r=1}^{L_r}&\alpha_G^{l_G}\alpha_r^{l_r}\\&\mathbf{a}_B\left(\varphi_G^{l_G}\right)\mathbf{a}_S^H\left(\vartheta_c^{l_G,l_r},\gamma_c^{l_G,l_r}\right),\end{aligned} \tag{7}$$

where $\vartheta_c^{l_G,l_r}=\vartheta_G^{l_G}-\vartheta_r^{l_r}$ and $\gamma_c^{l_G,l_r}=\gamma_G^{l_G}-\gamma_r^{l_r}$ denote the effective azimuth and elevation AoD spatial frequencies at the IRS.

B. Problem Formulation

We assume quasi-static block-fading channels, where $\mathbf{G}$, $\mathbf{h}_r$, and $\mathbf{h}_d$ remain constant during a certain period known as the coherence interval. For the sounding process, one coherence interval is divided into two stages, the first for channel estimation and the second for data transmission. For the general scenario with both the direct and cascaded channels, the channel estimation subinterval is further divided into $T+1$ blocks, where the first block and the other T blocks are used to estimate the direct and cascaded channels, respectively. For each channel estimation block, we consider a different reflection vector $\boldsymbol{\theta}_t=\left[s_{t,1}e^{j\phi_{t,1}},\ldots,s_{t,N_S}e^{j\phi_{t,N_S}}\right]^T$ for channel sensing, that is $\boldsymbol{\theta}_i\neq\boldsymbol{\theta}_j$ for all $0\leq i,j\leq T$ and $i\neq j$. During the training procedure of each block, the user transmits a known pilot x_t to the BS through the IRS with a reflecting vector $\boldsymbol{\theta}_t$. At the BS, the pilot is received as $\mathbf{y}_t\in\mathbb{C}^{Q\times1}$ through a (random) combining matrix $\mathbf{W}\in\mathbb{C}^{N_B\times Q}$. The received signal $\mathbf{y}_t$ for all the blocks ($t=0,\ldots,T$) assuming $x_t=1$ can be expressed as

$$\mathbf{y}_t=\mathbf{W}^H\left(\mathbf{H}_c\boldsymbol{\theta}_t+\mathbf{h}_d\right)+\mathbf{n}_t, \tag{8}$$

where $\mathbf{n}_t\sim\mathcal{CN}\left(\mathbf{0},\sigma_0^2\mathbf{I}_Q\right)$ denotes the additive white Gaussian noise vector with the noise power σ_0^2. Since we have only N_{RF}^B RF chains at the BS, we can only access a maximum N_{RF}^B-dimensional signal per symbol time. Hence, each pilot x_t needs to be sent $\lceil Q/N_{RF}^B\rceil$ times for the Q-dimensional signal $\mathbf{y}_t$ and the training overhead for channel estimation is $(T+1)\lceil Q/N_{RF}^B\rceil$.

During the first block for direct channel estimation ($t=0$), we assume all the IRS elements are turned off, i.e., $s_{0,n}=0$ for $n=1,\ldots,N_S$. Then, the received signal $\mathbf{y}_d\in\mathbb{C}^{Q\times1}$ becomes

$$\mathbf{y}_d=\mathbf{W}^H\mathbf{h}_d+\mathbf{n}_0. \tag{9}$$

For direct channel estimation, our goal is to estimate $\mathbf{h}_d$ from (9). Once $\mathbf{h}_d$ is estimated from (9), the effect of the direct channel can be removed from (8), and we can estimate the cascaded channel $\mathbf{H}_c$ accordingly.

Suppose the estimated direct channel is denoted as $\mathbf{h}_d^e$. During the following T blocks for cascaded channel estimation ($t=1,\ldots,T$), we assume all the IRS elements are turned on with random phases, i.e., $s_{t,n}=1$ and $\phi_{t,n}\sim\mathcal{U}\left(0,2\pi\right)$ for $n=1,\ldots,N_S$. By aggregating $\mathbf{y}_t$ for $t=1,\ldots,T$ as the overall measurement $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_T]\in\mathbb{C}^{Q\times T}$ for cascaded channel estimation, we have

$$\mathbf{Y}=\mathbf{W}^H\mathbf{H}_c\boldsymbol{\Theta}+\mathbf{W}^H\mathbf{h}_d\mathbf{1}_T+\mathbf{N}, \tag{10}$$

where $\boldsymbol{\Theta}=[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T]\in\mathbb{C}^{N_S\times T}$ is the reflection matrix for channel sensing, $\mathbf{1}_T=[1,\ldots,1]\in\mathbb{C}^{1\times T}$ is the one-vector of size T, and $\mathbf{N}=[\mathbf{n}_1,...,\mathbf{n}_T]\in\mathbb{C}^{Q\times T}$. After removing the effect of the direct channel, the overall received signal can be rewritten as

$$\mathbf{Y}_c=\mathbf{Y}-\mathbf{W}^H\mathbf{h}_d^e\mathbf{1}_T=\mathbf{W}^H\mathbf{H}_c\boldsymbol{\Theta}+\mathbf{W}^H\mathbf{e}_d\mathbf{1}_T+\mathbf{N}, \tag{11}$$

where $\mathbf{e}_d=\mathbf{h}_d-\mathbf{h}_d^e$ represent the estimation error of the direct channel. For cascaded channel estimation, our goal is to estimate $\mathbf{H}_c$ from (11), considering the estimation error of the direct channel for practical IRS systems.

To recover the direct and cascaded channels under low training overhead, we exploit the limited scattering nature of mmWave channels. Thus, the channel estimation problems can be formulated as sparse recovery problems. Considering the VAD representation, we can decompose $\mathbf{h}_d$ and $\mathbf{H}_c$ as [11]

$$\mathbf{h}_d\approx\mathbf{A}_B\widetilde{\mathbf{h}}_d, \tag{12}$$

$$\mathbf{H}_c\approx\mathbf{A}_B\widetilde{\mathbf{H}}_c\mathbf{A}_S^H, \tag{13}$$

where $\mathbf{A}_B\in\mathbb{C}^{N_B\times K_B}$ and $\mathbf{A}_S\in\mathbb{C}^{N_S\times K_S}$ are respectively the dictionary matrices at the BS and IRS, in which $K_B\geq N_B$ and $K_S\geq N_S$ are the number of grid points or the grid resolutions, while $K_S=K_S^hK_S^v$ with $K_S^h\geq N_S^h$ and $K_S^v\geq N_S^v$ being the horizontal and vertical resolutions of $\mathbf{A}_S$. Moreover, $\widetilde{\mathbf{h}}_d\in\mathbb{C}^{K_B\times1}$ and $\widetilde{\mathbf{H}}_c\in\mathbb{C}^{K_B\times K_S}$ are the angular direct and cascaded channels, which are a sparse vector and a row-sparse matrix, respectively. Here, (12) and (13) can be rewritten with equalities if the true angles fall perfectly on the grid points. Using (12) and (13), we can reformulate (9) and (11) as

$$\mathbf{y}_d\approx\mathbf{W}^H\mathbf{A}_B\widetilde{\mathbf{h}}_d+\mathbf{n}_0, \tag{14}$$

$$\mathbf{Y}_c\approx\mathbf{W}^H\mathbf{A}_B\widetilde{\mathbf{H}}_c\mathbf{A}_S^H\boldsymbol{\Theta}+\mathbf{W}^H\mathbf{e}_d\mathbf{1}_T+\mathbf{N}, \tag{15}$$

Therefore, known CS techniques, such as OMP and AMP, can be applied to recover the sparse $\widetilde{\mathbf{h}}_d$ and $\widetilde{\mathbf{H}}_c$ under low training overhead, i.e., $Q<N_B$ and $T<N_S$.

For the scenario without the direct channel due to blockage, we will focus on the challenging cascaded channel estimation problem. Since we don't have to estimate the unavailable direct channel, the channel estimation subinterval is only divided into T blocks for cascaded channel estimation. By assuming $\mathbf{h}_d=0$, we can express the overall measurement $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_T]\in\mathbb{C}^{Q\times T}$ as

$$\mathbf{Y}=\mathbf{W}^H\mathbf{H}_c\boldsymbol{\Theta}+\mathbf{N}\approx\mathbf{W}^H\mathbf{A}_B\widetilde{\mathbf{H}}_c\mathbf{A}_S^H\boldsymbol{\Theta}+\mathbf{N}, \tag{16}$$

Note that this scenario is equivalent to the case where the direct channel has been perfectly estimated in (15), i.e., $\mathbf{e}_d=0$.

C. Prior Arts: Compressive Channel Estimation (CCE)

Due to the large number of IRS elements without signal processing abilities, the training overhead for cascaded channel estimation is much larger than that for direct channel estimation. Hence, we will first focus on the challenging cascaded channel estimation problem, and then we will extend our system to the general system with both the direct and cascaded channels. By exploiting the sparsity nature of the angular cascaded channel, most works solve (16) with the following joint CS formulation [10]

$$\begin{aligned}\mathbf{y} = \text{vec}(\mathbf{Y}) &\approx \left(\mathbf{\Theta}^T \otimes \mathbf{W}^H\right) \text{vec}\left(\mathbf{A}_B \widetilde{\mathbf{H}}_c \mathbf{A}_S^H\right) + \mathbf{n} \\ &= \left(\mathbf{\Theta}^T \otimes \mathbf{W}^H\right)\left(\mathbf{A}_S^* \otimes \mathbf{A}_B\right)\widetilde{\mathbf{h}}_c + \mathbf{n} = \mathbf{A}\widetilde{\mathbf{h}}_c + \mathbf{n}, \quad (17)\end{aligned}$$

where $\widetilde{\mathbf{h}}_c = \text{vec}\left(\widetilde{\mathbf{H}}_c\right) \in \mathbb{C}^{K_B K_S \times 1}$, $\mathbf{n} = \text{vec}(\mathbf{N}) \in \mathbb{C}^{QT \times 1}$, and $\mathbf{A} = \left(\mathbf{\Theta}^T \otimes \mathbf{W}^H\right)\left(\mathbf{A}_S^* \otimes \mathbf{A}_B\right)$ is the measurement matrix of size $QT \times K_B K_S$.

In the joint CS method, the angular cascaded channel $\widetilde{\mathbf{h}}_c$ is directly recovered with the 3D measurement matrix $\mathbf{A}$ consisting of $K_B K_S$ atoms, where each atom corresponds to an AoA at the BS and a pair of effective azimuth and elevation AoDs at the IRS. The computational complexity of the joint CS method is $\mathcal{O}\left(L_G L_r T Q K_B K_S\right)$, which will be computationally prohibitive even in low grid resolutions, i.e., $K_B = N_B$ and $K_S = N_S$, due to the large numbers of BS antennas and IRS elements. Furthermore, since the number of cascaded paths $L_G L_r$ becomes larger due to the double scatters at the BS and the IRS, the sparsity of the angular cascaded channel is less significant compared with the standard MIMO angular channel. Hence, conventional CS algorithms, such as OMP and AMP, cannot achieve satisfactory estimation performance under low grid resolutions. Although the estimation accuracy can be improved with higher grid resolutions, there exists a trade-off between estimation accuracy and computational complexity.

To reduce the complexity of the joint CS method, [11] proposes a two-stage RIS-aided channel estimation (TRICE) framework to decouple the channel parameters. In the first stage, the AoA at the BS is estimated. In the second stage, the effective azimuth and elevation AoDs at the IRS are estimated. Since both stages are carried out through the estimation of the corresponding spatial frequencies, the estimation error mainly comes from the difference between the estimated spatial frequencies and the true spatial frequencies. However, since the performance improvement through increasing the grid resolutions is limited, the high grid resolutions used to achieve satisfactory estimation performance result in unaffordable computational complexity.

III. Proposed RCTS-LAMP Network for Cascaded Channel Estimation

Before going to the details of the proposed RCTS-LAMP network, we briefly review the LAMP network for the sparse signal recovery problem.

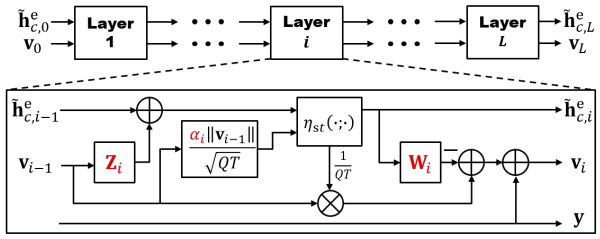


Fig. 2. The architecture and the i-th layer of the LAMP network.

A. LAMP Network

In contrast to the TRICE framework which requires the estimation of the true spatial frequencies, AMP is an iterative thresholding algorithm that directly recovers the sparse angular channel with the soft-thresholding function. Hence, we can recover each path of the cascaded channel with the grid points close to the true spatial frequencies. Although the AMP algorithm is effective for the general sparse recovery problem, the channel estimation performance of AMP is limited since the linear transform matrices and the shrinkage parameters are the same for all the iterations.

To boost the performance of AMP and make it suitable for the problem of cascaded channel estimation, we can unfold the iterations of the AMP algorithm into trainable layers through deep unfolding [29]. As illustrated in Fig. 2, the i-th layer outputs of a L-layer LAMP network are shown below

$$\widetilde{\mathbf{h}}_{c,i}^e = \eta_{st}\left(\widetilde{\mathbf{h}}_{c,i-1}^e + \mathbf{Z}_i \mathbf{v}_{i-1};\ \frac{\alpha_i}{\sqrt{QT}}\|\mathbf{v}_{i-1}\|\right), \quad (18)$$

$$\mathbf{v}_i = \mathbf{y} - \mathbf{W}_i \widetilde{\mathbf{h}}_{c,i}^e + \frac{1}{QT}\|\widetilde{\mathbf{h}}_{c,i}^e\|_0 \mathbf{v}_{i-1}, \quad (19)$$

where $\widetilde{\mathbf{h}}_{c,i}^e$ and $\mathbf{v}_i$ are respectively the estimated angular cascaded channel and the residual noise initialized as $\widetilde{\mathbf{h}}_{c,0}^e = \mathbf{0}$ and $\mathbf{v}_0 = \mathbf{y}$. The soft-thresholding function η_{st} not only provides the denoising function but also makes $\widetilde{\mathbf{h}}_{c,i}^e$ sparser by considering the sparsity. Specifically, the soft-thresholding function $\left[\eta_{st}\left(\mathbf{x};\lambda\right)\right]_j = \left(x_j - \lambda\frac{x_j}{|x_j|}\right)\mathbb{I}\left(|x_j| > \lambda\right)$ can shrink the elements of $\mathbf{x}$ whose amplitude is less than λ to zero with the indicator function $\mathbb{I}\left(\cdot\right)$. Besides, the term $\frac{1}{QT}\|\widetilde{\mathbf{h}}_{c,i}^e\|_0 \mathbf{v}_{i-1}$ is introduced for the calculation of $\mathbf{v}_i$ to accelerate convergence. To boost the cascaded channel estimation performance, we can optimize $\{\mathbf{W}_i, \mathbf{Z}_i, \alpha_i\}_{i=1}^L$ as the tunable parameters of the LAMP network through DL, where the linear transform matrices $\{\mathbf{W}_i\}_{i=1}^L$ and $\{\mathbf{Z}_i\}_{i=1}^L$ respectively replace the measurement matrix $\mathbf{A}$ and the matched filter $\mathbf{A}^H$ in the original AMP, and $\{\alpha_i\}_{i=1}^L$ are the shrinkage parameters for the soft-thresholding function. Finally, the cascaded channel is recovered by transforming the estimated angular cascaded channel $\widetilde{\mathbf{h}}_{c,L}^e$ into the time domain channel via $\mathbf{h}_{c,L}^e = \left(\mathbf{A}_S^* \otimes \mathbf{A}_B\right)\widetilde{\mathbf{h}}_{c,L}^e$.

However, the LAMP network directly recovers the cascaded channel by solving the joint recovery problem (17). The computational complexity of the LAMP network is $\mathcal{O}\left(LTQK_B K_S\right)$, which is computationally prohibitive for practical systems, even in low grid resolutions. To address this issue, in Section III-B, we propose a two-stage LAMP

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

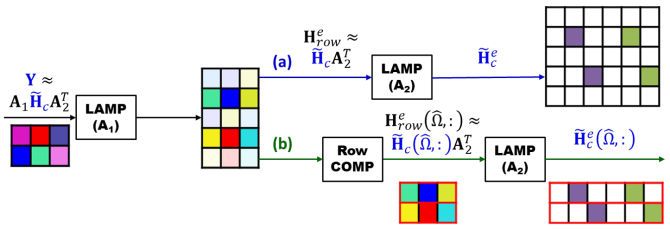


Fig. 3. The architecture of the proposed networks for the cascaded channel estimation in IRS-aided systems: (a) two-stage LAMP (TS-LAMP) network, and (b) TS-LAMP network with row compression (RCTS-LAMP).

network with row compression (RCTS-LAMP) for cascaded channel estimation, which will strike a better trade-off between estimation performance and computational complexity.

B. Two-Stage LAMP Network with Row Compression

Contrary to the joint CS method that directly recovers $\widetilde{\mathbf{H}}_c$ with a single LAMP network considering the 3D measurement matrix $\mathbf{A}$, the proposed two-stage LAMP (TS-LAMP) network recovers $\widetilde{\mathbf{H}}_c$ sequentially with two LAMP networks as illustrated in Fig. 3(a), where a row-sparse matrix $\mathbf{H}_{row}$ and $\widetilde{\mathbf{H}}_c$ are recovered in the first and second stages, respectively. Hence, the proposed TS-LAMP network can benefit from the high-performance LAMP network and enjoy low complexity by decomposing the channel recovery process. Moreover, by using the row-structured sparsity of the angular cascaded channel $\widetilde{\mathbf{H}}_c$, we introduce a row compression (COMP) mechanism to identify the row support $\hat{\Omega}$. Therefore, the proposed RCTS-LAMP network can further reduce the computational complexity by only recovering the rows of $\widetilde{\mathbf{H}}_c$ on the row support $\hat{\Omega}$, as illustrated in Fig. 3(b).

Instead of solving the joint recovery problem (17), the TS-LAMP network utilizes the structure resides in the overall measurement matrix $\mathbf{Y}$ of (16) as follows

$$\mathbf{Y} \approx \mathbf{A}_1 \widetilde{\mathbf{H}}_c \mathbf{A}_2^T + \mathbf{N}, \tag{20}$$

where $\mathbf{A}_1 = \mathbf{W}^H \mathbf{A}_B \in \mathbb{C}^{Q \times K_B}$ and $\mathbf{A}_2 = \mathbf{\Theta}^T \mathbf{A}_S^* \in \mathbb{C}^{T \times K_S}$ are the equivalent measurement matrices for $\widetilde{\mathbf{H}}_c$ at the BS and IRS, respectively. Furthermore, based on (7), each cascaded path (l_G, l_r) of the cascaded channel can provide a non-zero element for $\widetilde{\mathbf{H}}_c$, whose row and column index depend on $\varphi_G^{l_G}$ and $\left(\vartheta_c^{l_G,l_r}, \gamma_c^{l_G,l_r}\right)$, respectively. Therefore, the angular cascaded channel $\widetilde{\mathbf{H}}_c$ is row-structured. Namely, $\widetilde{\mathbf{H}}_c$ has L_G non-zero rows, and each non-zero row has L_r non-zero entries.

Based on the row structure of $\widetilde{\mathbf{H}}_c$, the proposed TS-LAMP network decomposes the cascaded channel estimation into two stages as summarized in Algorithm 1. In the first stage, we solve the following row-sparse matrix recovery problem

$$\mathbf{Y} = \mathbf{A}_1 \mathbf{H}_{row} + \mathbf{N}, \tag{21}$$

where $\mathbf{H}_{row} = \widetilde{\mathbf{H}}_c \mathbf{A}_2^T$ is a L_G row-sparse matrix since $\widetilde{\mathbf{H}}_c$ has L_G non-zero rows. Note that the recovery of $\mathbf{H}_{row}$ is an MMV problem, where each column of $\mathbf{H}_{row}$ results in a single measurement vector (SMV) in $\mathbf{Y}$. Since all the vectors in $\mathbf{Y}$ result from the common measure matrix $\mathbf{A}_1$, we can solve the MMV problem (21) with a single LAMP network through parallel computation. As the measurement matrix $\mathbf{A}$ in the joint CS method, we can generate the first LAMP network from the equivalent measure matrix $\mathbf{A}_1$ to recover the columns of $\mathbf{H}_{row}$. Through the parallel computation of all the SMVs in $\mathbf{Y}$, the columns of $\mathbf{H}_{row}$ can be recovered parallelly by the first LAMP network. The output of the first network is denoted as $\mathbf{H}_{row}^e$, which is the estimated value of $\mathbf{H}_{row}$.

In the second stage, we recover the angular cascaded channel $\widetilde{\mathbf{H}}_c$ from the L_G row-sparse matrix $\mathbf{H}_{row}^e$, which solves another MMV problem as follows

$$\left(\mathbf{H}_{row}^e\right)^T = \mathbf{A}_2 \widetilde{\mathbf{H}}_c^T + \mathbf{Z}, \tag{22}$$

where $\mathbf{Z}$ is the residual noise that stems from the estimation error of the first LAMP network. Since each column of $\widetilde{\mathbf{H}}_c^T$ results in a SMV in $\left(\mathbf{H}_{row}^e\right)^T$ from a common measurement matrix $\mathbf{A}_2$, the rows of $\widetilde{\mathbf{H}}_c$ can be recovered by another LAMP network. As in the first stage, we construct the second LAMP network from the equivalent measure matrix $\mathbf{A}_2$ to recover the rows of $\widetilde{\mathbf{H}}_c$ parallelly. The output of the second network is denoted as $\widetilde{\mathbf{H}}_c^e$, which is the estimated value of $\widetilde{\mathbf{H}}_c$. Finally, the estimated cascaded channel $\mathbf{H}_c^e$ is acquired by transforming the angular channel into the time domain with the dictionary matrices $\mathbf{A}_B$ and $\mathbf{A}_S$ as in Step 3 of Algorithm 1.

The proposed TS-LAMP network avoids the expensive 3D measurement matrix used in the joint CS method by decomposing the estimation process. Moreover, the two LAMP networks can be optimized through deep learning to achieve high performance. However, since the TS-LAMP network completely recovers all the $K_B K_S$ elements of the angular cascaded channel $\widetilde{\mathbf{H}}_c$ in the second stage, the computational complexity is still high under high grid resolutions K_B and K_S. Fortunately, due to the row-structured sparsity of the angular cascaded channel $\widetilde{\mathbf{H}}_c$, there exists redundancy in the second LAMP network. By focusing on the rows of the estimated row-sparse matrix whose norms are relatively higher, we can obtain a compact representation of the estimated row-sparse matrix and effectively reduce the computational complexity in the second stage.

The proposed RCTS-LAMP network is summarized in Algorithm 2, where the row compression mechanism is introduced between the two LAMP networks to obtain the row support $\hat{\Omega}$ of the estimated row-sparse matrix $\mathbf{H}_{row}^e$. The row support $\hat{\Omega}$ is used to sift out the useful rows of $\mathbf{H}_{row}^e$, which are expressed as $\mathbf{H}_{row}^e\left(\hat{\Omega}, :\right)$ and fed into the second LAMP network. Since the power level of a row $\mathbf{H}_{row}^e\left(i, :\right)$

Algorithm 1 Two-Stage LAMP (TS-LAMP) Network

Input: Overall measurement matrix $\mathbf{Y}$

1: **Stage 1:** Return the estimated row-sparse matrix $\mathbf{H}_{row}^e$ from $\mathbf{Y}$ with the LAMP network generated from $\mathbf{A}_1$.
2: **Stage 2:** Return the estimated angular cascaded channel $\widetilde{\mathbf{H}}_c^e$ from $\mathbf{H}_{row}^e$ with the LAMP network generated from $\mathbf{A}_2$.
3: $\mathbf{H}_c^e = \mathbf{A}_B \widetilde{\mathbf{H}}_c^e \mathbf{A}_S^H$
4: **return** Estimated cascaded channel $\mathbf{H}_c^e$

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

Algorithm 2 Two-Stage LAMP Network with Row Compression (RCTS-LAMP Network)

Input: Overall measurement matrix $\mathbf{Y}$

1: **Stage 1:** Return the estimated row-sparse matrix $\mathbf{H}_{row}^{e}$ from $\mathbf{Y}$ with the LAMP network generated from $\mathbf{A}_1$.
2: **Row Compression (COMP):** Return the row support $\hat{\Omega}$ of $\mathbf{H}_{row}^{e}$ by (23).
3: **Stage 2:** Return the estimated angular cascaded channel $\widetilde{\mathbf{H}}_{c}^{e}\left(\hat{\Omega},:\right)$ on the support $\hat{\Omega}$ from $\mathbf{H}_{row}^{e}\left(\hat{\Omega},:\right)$ with the LAMP network generated from $\mathbf{A}_2$.
4: $\mathbf{H}_{c}^{e}=\mathbf{A}_{B}\left(:,\hat{\Omega}\right)\widetilde{\mathbf{H}}_{c}^{e}\left(\hat{\Omega},:\right)\mathbf{A}_{S}^{H}$
5: **return** Estimated cascaded channel $\mathbf{H}_{c}^{e}$

will be large if the corresponding spatial frequency at the BS of that row is close to a true spatial frequency $\varphi_{G}^{l_G}$, we exploit the l_2-norm of a row $\|\mathbf{H}_{row}^{e}(i,:)\|$ to serve as an efficient power indicator. Since $\mathbf{H}_{row}^{e}$ is a L_G row-sparse matrix, the distribution of l_2-norms is positively skewed, where most of the rows have small l_2-norms and the few rows with large l_2-norms establish a long right tail of the distribution. Besides, since the rows of small norms mainly result from the residual noise, the mean of the norms will be close to the noise level, especially at high grid resolution K_B. Therefore, using the mean as the dynamic threshold is going to exclude the rows whose norms are less than or close to the noise level. Considering that the noise in the excluded rows can easily corrupt the information of the cascaded channel, we can neglect them in the second stage without performance loss, as shown in the simulation of Section V-C.

In contrast to the TS-LAMP network, the RCTS-LAMP network constructs a compact representation of $\mathbf{H}_{row}^{e}$ with row compression, where only the support rows $\hat{\Omega}$ are considered as the input to the second LAMP network. By considering the mean of l_2-norms as a dynamic threshold for row compression, we can filter out more than half of the rows since the mean is larger than the median in a positively skewed distribution. Specifically, the support rows $\hat{\Omega}$ are the row indices whose l_2-norms are larger than the dynamic threshold as follows

$$\hat{\Omega}=\left\{i:\|\mathbf{H}_{row}^{e}(i,:)\|>\sum_{j=1}^{K_B}\|\mathbf{H}_{row}^{e}(j,:)\|/K_B\right\},\quad(23)$$

where $\sum_{j=1}^{K_B}\|\mathbf{H}_{row}^{e}(j,:)\|/K_B$ denotes the dynamic threshold generated by the mean of all the l_2-norms. As a result, the second LAMP network only recovers $\widetilde{\mathbf{H}}_{c}^{e}\left(\hat{\Omega},:\right)$, where the other $K_B-\left|\hat{\Omega}\right|$ rows of $\widetilde{\mathbf{H}}_{c}^{e}$ are neglected for computation reduction and noise suppression. Finally, Step 4 of Algorithm 2 recovers the time domain channel $\mathbf{H}_{c}^{e}$ from the angular domain with the dictionary matrices $\mathbf{A}_{B}\left(:,\hat{\Omega}\right)$ and $\mathbf{A}_{S}$.

C. Learned Vector Shrinkage Function for the Recovery of the Row-Sparse Matrix

In Section III-B, the proposed TS-LAMP and RCTS-LAMP networks solve the MMV problems (21) and (22) from the SMV perspective. Although we can achieve high computational efficiency through parallel recovery of all the vectors with a common LAMP network, recovering the row-sparse matrix $\mathbf{H}_{row}$ from the MMV perspective may significantly improve the performance by exploiting the common sparsity. To solve the MMV problem (21) in a centralized approach, we consider vector AMP (VAMP) with the vector shrinkage function proposed in [30]. The VAMP starts with $\mathbf{H}_{row,0}^{e}=\mathbf{0}$ and $\mathbf{V}_0=\mathbf{Y}$, and the i-th iteration is given as follows

$$\mathbf{R}_i=\mathbf{H}_{row,i-1}^{e}+\mathbf{A}_1^{H}\mathbf{V}_{i-1},\quad(24)$$

$$\mathbf{h}_{row,i,j}^{e}=\boldsymbol{\eta}_{vst}\left(\mathbf{r}_{i,j};\ \lambda_i\right),\ j=1,\ldots,K_B\quad(25)$$

$$\mathbf{V}_i=\mathbf{Y}-\mathbf{A}_1\mathbf{H}_{row,i}^{e}+\frac{1}{Q}\mathbf{V}_{i-1}\sum_{j=1}^{K_B}\boldsymbol{\eta}_{vst}'\left(\mathbf{r}_{i,j};\ \lambda_i\right),\quad(26)$$

where $\boldsymbol{\eta}_{vst}:\mathbb{C}^T\rightarrow\mathbb{C}^T$ is the vector shrinkage function expressed as $\boldsymbol{\eta}_{vst}(\mathbf{x};\lambda)=\left(\mathbf{x}-\lambda\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right)\mathbb{I}\left(\|\mathbf{x}\|_2>\lambda\right)$ using the soft-thresholding function, $\lambda_i=\frac{\alpha}{\sqrt{Q}}\|\mathbf{V}_{i-1}\|$ is the shrinkage threshold, $\frac{1}{\sqrt{Q}}\|\mathbf{V}_{i-1}\|$ is the estimated standard deviation for each row in $\mathbf{V}_{i-1}$, $\mathbf{R}_i=\left[\mathbf{r}_{i,1}^{T},...,\mathbf{r}_{i,K_B}^{T}\right]^{T}$ is the input to the vector shrinkage function, $\mathbf{H}_{row,i}^{e}=\left[\left[\mathbf{h}_{row,i,1}^{e}\right]^{T},...,\left[\mathbf{h}_{row,i,K_B}^{e}\right]^{T}\right]^{T}$ is the estimated row-sparse matrix, and $\boldsymbol{\eta}_{vst}'(\mathbf{x};\ \lambda)=\frac{\partial\boldsymbol{\eta}_{vst}(\mathbf{x};\ \lambda)}{\partial\mathbf{x}}\in\mathbb{C}^{T\times T}$ is calculated for the Onsager correction of the VAMP algorithm.

Like the LAMP network which unfolds the iterations of AMP into trainable layers, we can unfold the VAMP algorithm into a LAMP-MMV network. The LAMP-MMV network utilizes the vector shrinkage function for MMV recovery and introduces tunable parameters $\{\mathbf{W}_i,\mathbf{Z}_i,\alpha_i\}_{i=1}^{L}$ into the VAMP algorithm. To be specific, the i-th layer of the LAMP-MMV network is shown as follows

$$\mathbf{R}_i=\mathbf{H}_{row,i-1}^{e}+\mathbf{Z}_i\mathbf{V}_{i-1},\quad(27)$$

$$\lambda_i=\frac{\alpha_i}{\sqrt{Q}}\|\mathbf{V}_{i-1}\|,\quad(28)$$

$$\mathbf{h}_{row,i,j}^{e}=\boldsymbol{\eta}_{vst}\left(\mathbf{r}_{i,j};\ \lambda_i\right),\ j=1,\ldots,K_B\quad(29)$$

$$\mathbf{V}_i=\mathbf{Y}-\mathbf{W}_i\mathbf{H}_{row,i}^{e}+\frac{1}{Q}\mathbf{V}_{i-1}\sum_{j=1}^{K_B}\boldsymbol{\eta}_{vst}'\left(\mathbf{r}_{i,j};\ \lambda_i\right),\quad(30)$$

where the tunable parameters $\{\mathbf{W}_i\}_{i=1}^{L}$, $\{\mathbf{Z}_i\}_{i=1}^{L}$, and $\{\alpha_i\}_{i=1}^{L}$ respectively replace the fixed parameters $\mathbf{A}_1$, $\mathbf{A}_1^{H}$, and α in the VAMP algorithm.

Since the LAMP-MMV network adopts the vector shrinkage function, each row $\mathbf{h}_{row,j}^{e}$ in the row-sparse matrix $\mathbf{H}_{row}^{e}$ can be recovered collectively to exploit the common sparsity in $\mathbf{H}_{row}^{e}$. Furthermore, the tunable parameters of LAMP-MMV can be optimized to have different values in different layers through deep unfolding, which can improve the denoising ability of the vector shrinkage function and improve the recovery performance of the row-sparse matrix $\mathbf{H}_{row}^{e}$. Compared with the LAMP network, the LAMP-MMV network has the same tunable parameters $\{\mathbf{W}_i,\mathbf{Z}_i,\alpha_i\}_{i=1}^{L}$. Hence, we can replace the first LAMP network in TS-LAMP and RCTS-LAMP networks with the LAMP-MMV network to improve

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

Algorithm 3 RCTS-LAMP Network for MMV Recovery (RCTS-LAMP-MMV Network)

Input: Overall measurement matrix $\mathbf{Y}$

1: **Stage 1:** Return the estimated row-sparse matrix $\mathbf{H}_{row}^{e}$ from $\mathbf{Y}$ with the LAMP-MMV network generated from $\mathbf{A}_1$ by (27)-(29).
2: **Row Compression (COMP):** Return the row support $\hat{\Omega}$ of $\mathbf{H}_{row}^{e}$ by (23).
3: **Stage 2:** Return the estimated angular cascaded channel $\widetilde{\mathbf{H}}_{c}^{e}\left(\hat{\Omega},:\right)$ on the support $\hat{\Omega}$ from $\mathbf{H}_{row}^{e}\left(\hat{\Omega},:\right)$ with the LAMP network generated from $\mathbf{A}_2$.
4: $\mathbf{H}_{c}^{e}=\mathbf{A}_{B}\left(:,\hat{\Omega}\right)\widetilde{\mathbf{H}}_{c}^{e}\left(\hat{\Omega},:\right)\mathbf{A}_{S}^{H}$
5: **return** Estimated cascaded channel $\mathbf{H}_{c}^{e}$

the performance of cascaded channel estimation without much complexity increase. We outline the RCTS-LAMP network for MMV recovery (RCTS-LAMP-MMV) in Algorithm 3, where the LAMP-MMV network is used in the first stage for better recovery performance. Besides, the case considering the TS-LAMP network for MMV recovery, i.e., TS-LAMP-MMV, can be derived similarly.

IV. Joint Estimation of Direct and Cascaded Channels via the Proposed RCTS-LAMP Network

Although the proposed RCTS-LAMP network focuses on the challenging cascaded channel estimation problem assuming that the direct channel is blocked, we can extend our system to consider the joint estimation problem of the direct and cascaded channels without modifying the network architecture. Observe that the direct channel recovery problem (14) is a sparse recovery problem as follows

$$\mathbf{y}_d \approx \mathbf{A}_1\widetilde{\mathbf{h}}_d + \mathbf{n}_0, \tag{31}$$

where the angular direct channel $\widetilde{\mathbf{h}}_d$ is measured by the same matrix $\mathbf{A}_1 = \mathbf{W}^H\mathbf{A}_B \in \mathbb{C}^{Q\times K_B}$ as the row-sparse matrix $\mathbf{H}_{row}$ in the row-sparse matrix recovery problem (21). Hence, we can construct a LAMP network from the equivalent measure matrix $\mathbf{A}_1$ to recover the angular direct channel $\widetilde{\mathbf{h}}_d$ with high estimation accuracy. After the estimation of the direct channel, the RCTS-LAMP network can be used to solve the cascaded channel estimation problem (15) considering the estimation error of the direct channel $\mathbf{e}_d$. However, the extra LAMP network for direct channel estimation results in additional memory and training overhead, which may hinder the realizability of the networks for practical IRS-aided communication systems.

Although two individual LAMP networks for the estimation of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ may increase the representation power and achieve better estimation performance, the training and memory overhead for a new LAMP network may be time-consuming and redundant. Since the row-sparse matrix $\mathbf{H}_{row}$ and the angular direct channel $\widetilde{\mathbf{h}}_d$ are both measured by the same equivalent matrix $\mathbf{A}_1$, we propose to use the first LAMP network in the RCTS-LAMP network for both the estimation of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$. Therefore, the knowledge learned from the estimation of $\mathbf{H}_{row}$ could be applied to the estimation of $\widetilde{\mathbf{h}}_d$. To make the first LAMP network multifunctional and optimize the RCTS-LAMP network under the estimation error $\mathbf{e}_d$, we propose a three-stage training procedure for the RCTS-LAMP network, as shown in Fig. 4. Note that the three-stage training procedure is also valid for the RCTS-LAMP-MMV network since the LAMP-MMV network will degenerate into the LAMP network when there is only one measurement vector. Hence, we can use the LAMP-MMV for both the estimation of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$. However, for ease of exposition, we elaborate on the three-stage training procedure for the RCTS-LAMP network.

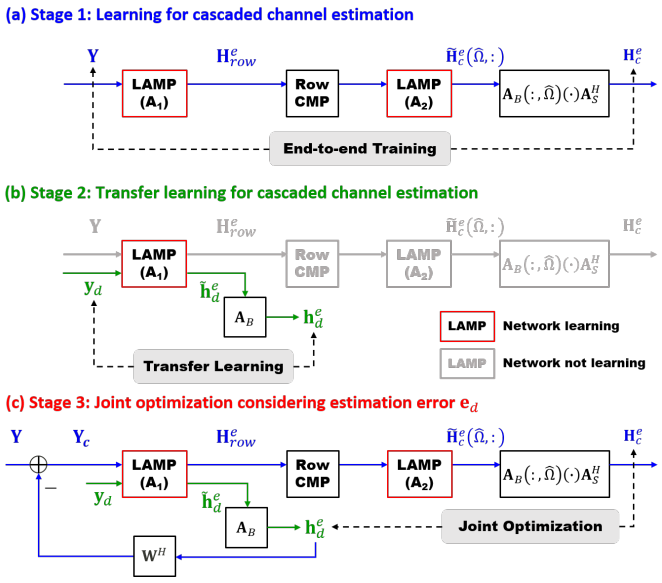


Fig. 4. Proposed three-stage training procedure for the joint optimization of the direct and cascaded channel via the RCTS-LAMP network.

In the first stage of the training procedure, we consider the cascaded channel estimation problem under the scenario where the direct channel is blocked or has been perfectly estimated (16). To recover the cascaded channel $\mathbf{H}_c$, the RCTS-LAMP network accepts the overall received signal $\mathbf{Y}$ as input and outputs the estimated cascaded channel $\mathbf{H}_c^e$ through the two LAMP networks, as shown in Fig. 4(a). To train the network, the input-output pair $(\mathbf{Y},\mathbf{H}_c)$ can be generated from (7) and (16). The training data $\mathcal{D}_{CC} = \left\{\left(\mathbf{Y}^{(u)},\mathbf{H}_c^{(u)}\right)\right\}_{u=1}^{u_{CC}}$ can be obtained by generating the input-output pairs for u_{CC} realizations. Before optimizing the tunable parameters $\mathcal{W}_1=\{\mathbf{W}_{i,1},\mathbf{Z}_{i,1},\alpha_{i,1}\}_{i=1}^{L}$ and $\mathcal{W}_2=\{\mathbf{W}_{i,2},\mathbf{Z}_{i,2},\alpha_{i,2}\}_{i=1}^{L}$ of the two LAMP networks, we initialize $\mathcal{W}_1$ and $\mathcal{W}_2$ as $\mathbf{W}_{i,1}=\mathbf{Z}_{i,1}^H=\mathbf{A}_1$, $\mathbf{W}_{i,2}=\mathbf{Z}_{i,2}^H=\mathbf{A}_2$ and $\alpha_{i,1}=\alpha_{i,2}=constant$ as the original AMP algorithm, which serves as an effective initial point with performance guarantees. Afterward, the tunable parameters $\mathcal{W}_1$ and $\mathcal{W}_2$ can be optimized by end-to-end training to minimize the normalized mean square error (NMSE) loss of $\mathbf{H}_c$ defined as follows

$$\mathcal{L}_{CC} = \frac{1}{u_{CC}}\sum_{u=1}^{u_{CC}}\left\{\|\mathbf{H}_c^{(u)}-\mathbf{H}_c^{e,(u)}\|^2/\|\mathbf{H}_c^{(u)}\|^2\right\}, \tag{32}$$

where $\mathbf{H}_c^{e,(u)}$ denotes the estimated cascaded channel for each realization u. After the first stage of training, the RCTS-LAMP

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

network can be utilized to estimate the cascaded channel when the direct channel is blocked or has been perfectly estimated. Note that this scenario is equivalent to the general system without the estimation error of the direct channel, and we consider this scenario as a stepping stone for the general system where both the direct and cascaded channels exist.

In the second stage of the training procedure, we consider the direct channel estimation problem (31). Although the first LAMP network has learned how to recover a row-sparse signal $\mathbf{H}_{row}^{e}$ from the measurement of $\mathbf{A}_1$, the loss function in the first stage is the NMSE loss of the $\mathbf{H}_c$, which only makes the two LAMP networks suitable for cascaded channel estimation. To make the first LAMP network capable of direct channel estimation, we apply transfer learning to the first LAMP network, as illustrated in Fig. 4(b). Hence, the knowledge gained from the recovery of the row-sparse signal $\mathbf{H}_{row}^{e}$ could be applied to the recovery of the sparse $\widetilde{\mathbf{h}}_d$. To recover the direct channel $\mathbf{h}_d$, the first LAMP network accepts the received signal $\mathbf{y}_d$ from (9) and output the angular direct channel $\widetilde{\mathbf{h}}_d^e$, which is transformed into the estimated direct channel $\mathbf{h}_d^e$ in the time domain via $\mathbf{h}_d^e = \mathbf{A}_B\widetilde{\mathbf{h}}_d^e$. The training data $\mathcal{D}_{DC} = \left\{\left(\mathbf{y}_d^{(u)}, \mathbf{h}_d^{(u)}\right)\right\}_{u=1}^{u_{DC}}$ can be generated via (3) and (9) with u_{DC} realizations. Based on the parameters $\mathcal{W}_1$ learned from stage 1, we fine-tune the parameters $\mathcal{W}_1$ by transfer learning to minimize the NMSE loss of $\mathbf{h}_d$ defined as follows

$$\mathcal{L}_{DC} = \frac{1}{u_{DC}} \sum_{u=1}^{u_{DC}} \left\{ \|\mathbf{h}_d^{(u)} - \mathbf{h}_d^{e,(u)}\|^2 / \|\mathbf{h}_d^{(u)}\|^2 \right\}, \tag{33}$$

where $\mathbf{h}_d^{e,(u)}$ denotes the estimated direct channel for each realization u. After the second stage of training, the first LAMP network is capable of recovering $\mathbf{h}_d$, which can be utilized to solve the direct channel estimation problem (31) for the general system with the direct link.

In the third stage of the training procedure, we consider the general system where both the direct and reflected links exist. Since the RCTS-LAMP network has sequentially learned how to estimate the cascaded and direct channels in the first two stages of training, the first and second LAMP networks can recover the sparse signals measured from the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, respectively. However, to estimate the cascaded channel under the existence of the direct channel (10), the direct channel's effect must be removed beforehand. Hence, in the third stage of training, the RCTS-LAMP network will learn how to jointly solve the direct channel estimation problem (14) and the cascaded channel estimation problem (15), considering the estimation error of the direct channel, as shown in Fig. 4(c). To estimate the direct channel, the RCTS-LAMP network accepts the received signal $\mathbf{y}_d$ from (9) and outputs the estimated direct channel $\mathbf{h}_d^e$ through the first LAMP network. The estimated direct channel $\mathbf{h}_d^e$ is used to cancel out the effect of the direct channel in $\mathbf{Y}$ from (10), which is converted to the cascaded channel estimation problem (15). From the measurement $\mathbf{Y}_c = \mathbf{Y} - \mathbf{W}^H\mathbf{h}_d^e\mathbf{1}_T$ in (11), the RCTS-LAMP network will output the estimated cascaded channel $\mathbf{H}_c^e$ through its two LAMP networks. The training data is generated via (3), (9), (7) and (10) as $\mathcal{D}_{JO} = \left\{\left(\mathbf{y}_d^{(u)}, \mathbf{h}_d^{(u)}\right), \left(\mathbf{Y}^{(u)}, \mathbf{H}_c^{(u)}\right)\right\}_{u=1}^{u_{JO}}$,

TABLE I
COMPARISON OF WEIGHT-SHARING METHODS FOR THE FIRST LAMP NETWORK

Method	Parameter	Memory Overhead	Capability
All shared	$\mathcal{W}_1 = \{\mathbf{W}_{i,1}, \mathbf{Z}_{i,1}, \alpha_{i,1}\}_{i=1}^{L}$	Low	Low
None Shared	$\mathcal{W}_1 = \{\mathbf{W}_{i,1}, \mathbf{Z}_{i,1}, \alpha_{i,1}\}_{i=1}^{L}$ $\mathcal{W}_d = \{\mathbf{W}_{i,d}, \mathbf{Z}_{i,d}, \alpha_{i,d}\}_{i=1}^{L}$	High	High
Separate α	$\mathcal{W}_1 = \{\mathbf{W}_{i,1}, \mathbf{Z}_{i,1}, \alpha_{i,1}\}_{i=1}^{L}$ $\mathcal{W}_d = \{\alpha_{i,d}\}_{i=1}^{L}$	Low	Medium

where each realization u consists of two input-output pairs for the two estimation problems. To jointly optimize the tunable parameters $\mathcal{W}_1$ and $\mathcal{W}_2$ of the LAMP networks, we consider the joint NMSE loss for both $\mathbf{h}_d$ and $\mathbf{H}_c$ defined as follows

$$\mathcal{L}_{JO} = \frac{1}{u_{JO}} \sum_{u=1}^{u_{JO}} \left\{ \frac{\|\mathbf{h}_d^{(u)} - \mathbf{h}_d^{e,(u)}\|^2}{\|\mathbf{h}_d^{(u)}\|^2} + \frac{\|\mathbf{H}_c^{(u)} - \mathbf{H}_c^{e,(u)}\|^2}{\|\mathbf{H}_c^{(u)}\|^2} \right\}, \tag{34}$$

where $\mathbf{h}_d^{e,(u)}$ and $\mathbf{H}_c^{e,(u)}$ are respectively the estimated direct and cascaded channels for each realization u. Therefore, we can further fine-tune the parameters $\mathcal{W}_1$ and $\mathcal{W}_2$, which will make the two LAMP networks capable of the recovery of $\mathbf{H}_c$ under the influence of the estimation error generated from the estimation of $\mathbf{h}_d$ through the first LAMP network.

Although the three-stage training procedure is designed to make the first LAMP network multifunctional with a shared tunable parameter $\mathcal{W}_1 = \{\mathbf{W}_{i,1}, \mathbf{Z}_{i,1}, \alpha_{i,1}\}_{i=1}^{L}$, this procedure is also applicable to the methods with additional tunable parameters $\mathcal{W}_d$ for the first LAMP network during direct channel estimation. That is, we need to consider the tunable parameter $\mathcal{W}_d$ during direct channel estimation in stages 2 and 3 of the training procedures. In stage 2, $\mathcal{W}_d$ is initialized by the parameter $\mathcal{W}_1$ learned from stage 1 and optimized via transfer learning. In stage 3, $\mathcal{W}_d$ is jointly optimized with $\mathcal{W}_1$ and $\mathcal{W}_2$ considering the joint NMSE loss $\mathcal{L}_{JO}$ in (34).

Although the additional tunable parameter $\mathcal{W}_d$ may increase the representative power of the first LAMP network and improve the estimation performance of the direct and cascaded channels, the additional memory overhead may hinder the deployment in practical systems. We consider various weight-sharing methods to achieve a better trade-off between estimation performance and memory overhead, as summarized in Table I. The two extreme cases are the all-shared method and the none-shared method. The all-shared method will make the first LAMP network of RCTS-LAMP capable of both direct and cascaded channel estimations. Besides, the none-shared method will equivalently create an extra LAMP network for direct channel estimation.

Considering that the linear transform matrices ($\mathbf{W}_i$ and $\mathbf{Z}_i$) and the shrinkage parameters (α_i) have different functions in the LAMP network, we also take the method of separate α into account. The method of separate α incurs negligible memory overhead while it can provide customized shrinkage parameters for the recovery of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ by considering their individual signal power. On the other hand, since both $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ are measured by the same matrix $\mathbf{A}_1$ and they

have similar signal patterns, we can use the same transform matrices $\mathbf{W}_i$ and $\mathbf{Z}_i$ for the recovery of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ without performance loss. As a result, the method of separate α is more suitable for the recovery of $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ since it can improve the estimation performance with negligible memory overhead.

V. Performance Evaluation

In this section, we evaluate the performance of cascaded channel estimation for the proposed RCTS-LAMP network and RCTS-LAMP-MMV network under the scenario where the direct channel is blocked or has been perfectly estimated. We consider the scenario where both direct and indirect links exist to validate the effectiveness of the three-stage training procedure for the RCTS-LAMP network. We provide several state-of-the-art benchmarks for comparison. Besides, we detail the simulation setting as well as performance metrics. Finally, the simulation results are provided with in-depth analysis and discussion.

A. Benchmarks

In our simulation, we compare the proposed RCTS-LAMP and RCTS-LAMP-MMV networks with the following benchmarks for cascaded channel estimation:

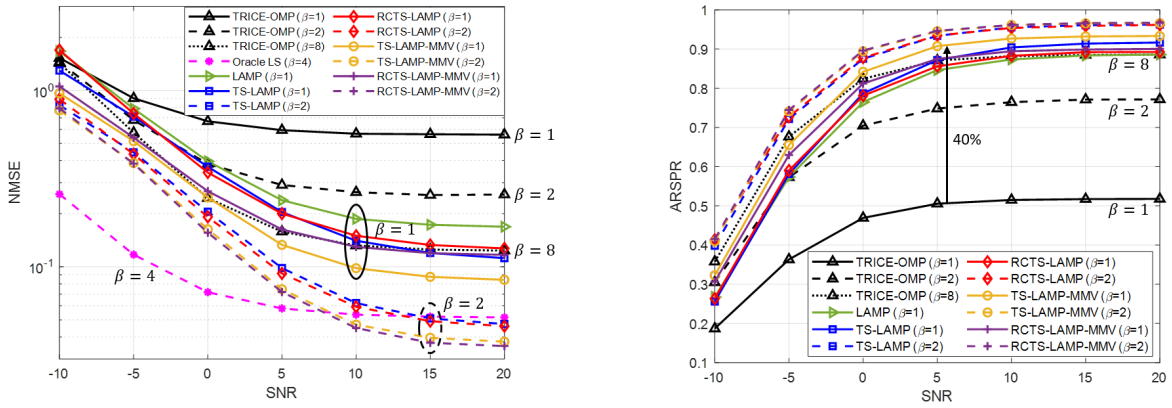
- *Benchmark 1. TRICE framework with OMP method (TRICE-OMP):* The TRICE framework solves the cascaded channel estimation problem (16) in two stages. In the first stage, the L_G AoAs at the BS are estimated by recovering the L_G row-sparse matrix $\mathbf{H}_{row}$ in (21) with the on-grid OMP method. Based on each estimated AoA at the BS $\hat{\varphi}_G^{l_G}$, we can again use the OMP method to recover the L_r tuples $\left(\vartheta_c^{l_G,l_r}, \gamma_c^{l_G,l_r}, \alpha_G^{l_G}\alpha_r^{l_r}\right)$, consisting of the effective azimuth and elevation AoDs at the IRS and the effective complex path gains. Finally, the estimated cascaded channel can be recovered by replacing the true channel parameters in (7) with the estimated channel parameters.
- *Benchmark 2. On-grid recovery with perfectly known supports (Oracle LS):* Oracle LS can be seen as the performance upper bound for on-grid CS methods. Note that the performance of the on-grid OMP method relies on the grid resolutions, which quantizes the AoA at the BS and effective AoDs at the IRS into K_B and K_S levels, respectively. For each cascaded path (l_G, l_r) of the cascaded channel, the support of this path corresponds to one of the $K_B K_S$ grid points that is closest to the true channel parameters. Oracle LS is a genie-aided scheme, where the supports Ω_O of the vectorized angular channel $\widetilde{\mathbf{h}}_c$ are known perfectly. Hence, the minimum variance unbiased estimator for (17) is the least squares (LS) estimator as follows $\widetilde{\mathbf{h}}_c^e(\Omega_O) = \mathbf{A}(:, \Omega_O)^\dagger \mathbf{y}$.
- *Benchmark 3. Joint-CS recovery with LAMP network (LAMP):* To understand the benefit of the proposed RCTS-LAMP network that decomposes the cascaded channel estimation into two stages, we implement the LAMP network to solve the joint CS problem (17) directly. The LAMP learns how to recover the vectorized angular cascaded channel $\widetilde{\mathbf{h}}_c$ through end-to-end training, where the LAMP network is constructed with the 3D measurement matrix $\mathbf{A}$ in (17). Although the estimation performance can be increased through end-to-end training, the computational complexity is high due to the large matrix size of the 3D measurement matrix $\mathbf{A}$.

B. Simulation Setting and Performance Metrics

In our simulation, the system parameters are set as follows: the number of BS antennas $N_B = 32$, the number of IRS elements $N_S = 64$ $\left(N_S^h = N_S^v = 8\right)$, the number of measurements at the BS $Q = 24$, and the number of paths of the channels $L_G = L_r = L_d = 3$. For both training and testing, the angular-domain path angles (φ_a, ϑ_a and γ_a) of the channels are uniformly distributed as $\mathcal{U}(-\pi/2, \pi/2)$, which may not lie on the pre-discretized virtual angle grids of the dictionary matrices $\mathbf{A}_B$ and $\mathbf{A}_S$. Therefore, the proposed RCTS-LAMP networks can learn to estimate the direct and cascaded channels under general channel parameters. The training procedure is a one-time cost and overhead since the RCTS-LAMP networks can generalize to various channel parameters and do not require retraining. We adopt practical path gains in IRS-aided systems. Namely, $\alpha_G^{l_G} \sim \mathcal{CN}\left(0, 10^{-3} d_G^{-2.2}\right)$, $\alpha_r^{l_r} \sim \mathcal{CN}\left(0, 10^{-3} d_r^{-2.8}\right)$, and $\alpha_d^{l_d} \sim \mathcal{CN}\left(0, 10^{-4} d_d^{-3.5}\right)$, where the distance between the BS and IRS, the distance between the IRS and user, and the distance between the BS and user are assumed to be $d_G = 10m$, $d_r = 100m$, and $d_d = 100m$, respectively. The SNR is defined as $\mathbb{E}\left\{\|\mathbf{W}^H \mathbf{H}_c \mathbf{\Theta}\|^2 / \|\mathbf{N}\|^2\right\}$ for cascaded channel estimation, where the elements of $\mathbf{W}$ and $\mathbf{\Theta}$ have unit gains and random phases following the $\mathcal{U}(0, 2\pi)$ distribution. Based on the above system parameters, we generate 10^5 independent samples with SNR $= 5$dB for each stage of the training procedure, i.e., $u_{CC} = u_{DC} = u_{JO} = 10^5$. To evaluate the estimation performance, we generate 10^4 independent samples for each testing SNR(dB)$\in \{-10, -5, 0, 5, 10, 15, 20\}$. We use the deep learning library PyTorch for the training and testing of the proposed networks. During training, the Adam optimizer is adopted with a learning rate of 10^{-4}.

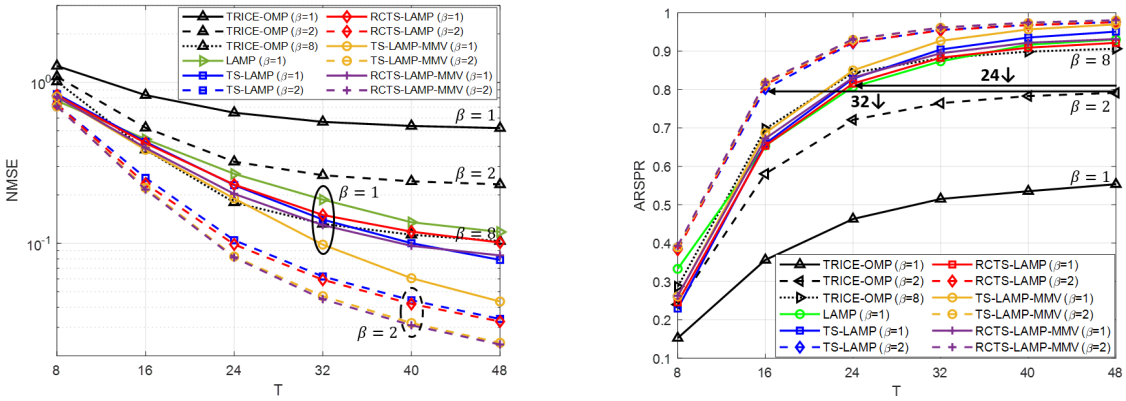
To evaluate the performance under different grid resolutions, we define the oversampling rate $\beta = \frac{K_B}{N_B} = \frac{K_S^h}{N_S^h} = \frac{K_S^v}{N_S^v}$ to indicate the grid resolutions. To evaluate the channel estimation performance, we define the NMSE metrics as $\mathbb{E}\left\{\|\mathbf{H}_c - \mathbf{H}_c^e\|^2 / \|\mathbf{H}_c\|^2\right\}$ and $\mathbb{E}\left\{\|\mathbf{h}_d - \mathbf{h}_d^e\|^2 / \|\mathbf{h}_d\|^2\right\}$ for cascaded and direct channel estimation, respectively. Since the NMSE performance cannot directly reflect the quality of communication, we adopt the metric of average receive signal power ratio (ARSPR) [10] to validate the effectiveness of the proposed methods with respect to communications. ARSPR is defined as the ratio of the actual receive signal power to the ideal receive signal power. Specifically, ARSPR is given as $\mathbb{E}\left\{\frac{\left|\left(\mathbf{h}_r^H \mathrm{diag}(\hat{\mathbf{v}}) \mathbf{G}^H + \mathbf{h}_d^H\right)\hat{\mathbf{f}}\right|^2}{\left|\left(\mathbf{h}_r^H \mathrm{diag}(\mathbf{v}) \mathbf{G}^H + \mathbf{h}_d^H\right)\mathbf{f}\right|^2}\right\}$, where $\left(\hat{\mathbf{v}}, \hat{\mathbf{f}}\right)$ and $(\mathbf{v}, \mathbf{f})$ are obtained by solving the joint beamforming problem with the fixed point iteration (FPI) algorithm [31] based on the estimated channels $(\mathbf{H}_c^e, \mathbf{h}_d^e)$ and the true channels $(\mathbf{H}_c, \mathbf{h}_d)$, respectively.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845



(a) (b)

Fig. 5. Comparison of cascaded channel estimation performance with respect to different performance metrics versus SNR under $T = 32$ channel estimation blocks: (a) NMSE versus SNR, (b) ARSPR versus SNR.



(a) (b)

Fig. 6. Comparison of cascaded channel estimation performance with respect to different performance metrics versus channel estimation blocks T under SNR $= 10$ dB: (a) NMSE versus T, (b) ARSPR versus T.

C. Performance of Cascaded Channel Estimation under Perfect Direct Channel

In this subsection, we consider the scenario where the direct channel is blocked and focus on the challenging cascaded channel estimation problem. As shown in Fig. 5, we compare the cascaded channel estimation performance with respect to the performance metrics of NMSE and ARSPR under $T = 32$ channel estimation blocks. Fig. 5(a) shows the NMSE performance comparison versus SNR. The performance of TRICE-OMP is simulated under $\beta = 1$, 2, and 8. We can observe that TRICE-OMP has an NMSE of 0.5 under SNR $= 5$ and $\beta = 1$, which shows unsatisfactory performance under low grid resolutions. Although the performance can be improved with higher β of 2 and 8, the computational complexity will increase drastically to achieve a satisfactory estimation performance. Besides, the LAMP, TS-LAMP, and RCTS-LAMP networks with $\beta = 1$ can outperform TRICE-OMP with $\beta = 2$, demonstrating the effectiveness of DL methods compared to on-grid CS methods by learning effective linear transform matrices and shrinkage parameters for channel recovery.

Instead of solving the high complexity joint CS problem as LAMP, the proposed TS-LAMP and RCTS-LAMP networks enjoy lower computational complexity with the two-stage networks while achieving better NMSE performance than the LAMP network. Also, since the proposed TS-LAMP and RCTS-LAMP networks have almost the same performance, the additional row compression block in the RCTS-LAMP network effectively reduces the complexity without performance loss. Furthermore, the proposed TS-LAMP and RCTS-LAMP networks with $\beta = 2$ significantly increase the estimation performance, which outperforms TRICE-OMP with $\beta = 8$ and achieves Oracle LS with $\beta = 4$ at the high SNR region.

By exploiting the common sparsity of the row-sparse matrix $\mathbf{H}^e_{row}$, the proposed TS-LAMP-MMV and RCTS-LAMP-MMV networks achieve improved cascaded channel estimation performance with the learned vector shrinkage function. Although the RCTS-LAMP-MMV may have degraded performance under $\beta = 1$ due to row compression, both the TS-LAMP-MMV and RCTS-LAMP-MMV networks with $\beta = 1$ can outperform TRICE-OMP with $\beta = 8$. However, TS-LAMP-MMV and RCTS-LAMP-MMV networks with $\beta = 2$ have almost the same performance, which outperforms Oracle LS with $\beta = 4$ at the high SNR region.

To reflect the quality of communications of the proposed methods, we simulate the ARSPR performance as shown in Fig. 5(b). As expected, the lower NMSE performance implies that the corresponding method can achieve a better ARSPR performance. Under the same $\beta = 1$, the LAMP-

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845

TABLE II
COMPARISON OF COMPUTATIONAL COMPLEXITY

Complexity	Computational Complexity	Execution Time (ms)
TRICE-OMP	$\mathcal{O}\left(L_G TQK_B + L_r TK_S\right)$	2.9 ($\beta = 1$) 3.2 ($\beta = 2$) 9.5 ($\beta = 8$)
LAMP	$\mathcal{O}\left(LTQK_BK_S\right)$	24.5 ($\beta = 1, L = 10$)
TS-LAMP	$\mathcal{O}\left(LTQK_B + LTK_BK_S\right)$	4.2 ($\beta = 1, L = 10$) 3.4 ($\beta = 2, L = 6$)
RCTS-LAMP	$\mathcal{O}\left(LTQK_B + LT\left\|\hat{\Omega}\right\|K_S\right)$	2.5 ($\beta = 1, L = 10$) 1.8 ($\beta = 2, L = 6$)
TS-LAMP-MMV	$\mathcal{O}\left(LTQK_B + L(Q+K_B)T^2 + LTK_BK_S\right)$	5.8 ($\beta = 1, L = 10$) 4.3 ($\beta = 2, L = 6$)
RCTS-LAMP-MMV	$\mathcal{O}\left(LTQK_B + L(Q+K_B)T^2 + LT\left\|\hat{\Omega}\right\|K_S\right)$	4.1 ($\beta = 1, L = 10$) 2.9 ($\beta = 2, L = 6$)

based networks can increase the ARSPR by 35%-40% at SNR $= 5$ compared with TRICE-OMP. Moreover, for the proposed (RC)TS-LAMP and (RC)TS-LAMP-MMV networks with $\beta = 2$, the ARSPR performance is larger than 95% for SNR larger than 10 dB, which demonstrates the effectiveness with respect to communications.

To evaluate the cascaded channel estimation performance under low training overhead, we compare the NMSEs and APRSRs of respective methods versus channel estimation blocks T as shown in Fig. 6. Note that the training overhead for cascaded channel estimation is $T\left\lceil Q/N_{RF}^B\right\rceil$, which is proportional to the channel estimation blocks T. From Fig. 6(a), we can observe that the NMSE decreases with the increase of T. Moreover, the LAMP, (RC)TS-LAMP, and (RC)TS-LAMP-MMV networks with $\beta = 1$ can achieve TRICE-OMP with $\beta = 8$, which implies that the cascaded channel can be recovered under low grid resolutions and low pilot overhead. To reflect the quality of communications with respect to the training overhead, Fig. 6(b) shows the ARSPR performance versus T. From Fig. 6(b), we can observe that the proposed (RC)TS-LAMP and (RC)TS-LAMP-MMV networks require much lower training overhead than TRICE-OMP. Specifically, compared with TRICE-OMP ($\beta = 2$), the proposed (RC)TS-LAMP and (RC)TS-LAMP-MMV networks can reduce T by 24 and 32 under the oversampling rates β of 1 and 2, respectively.

Table II summarizes the comparison of computational complexity and execution time. Note that since the computation for each method is the same for any SNR, the comparison in Table II is independent of SNR. The execution time is simulated on an AMD Ryzen Threadripper PRO 3995WX CPU under the same system parameters as Fig. 5. Although the TRICE-OMP enjoys low computational complexity by decoupling the channel parameters, the execution time increases with the grid resolutions. As shown in Fig. 5, to achieve the performance of LAMP-based networks at $\beta = 1$, TRICE-OMP requires high grid resolutions of $\beta = 8$ with an execution time of 9.5 ms. The computational complexity of the LAMP network is dominated by the multiplications of the linear transform matrices, which is $\mathcal{O}\left(TQK_BK_S\right)$ in each iteration. Compared with the LAMP network, the TS-LAMP network enjoys lower computational complexity $\mathcal{O}\left(LTQK_B + LTK_BK_S\right)$ by constructing two small LAMP networks with the equivalent measurement matrices $\mathbf{A}_1 \in \mathbb{C}^{Q\times K_B}$ and $\mathbf{A}_2 \in \mathbb{C}^{T\times K_S}$, which correspond to the linear transformations of the T SMVs in the first stage and K_B SMVs in the second stage, respectively.

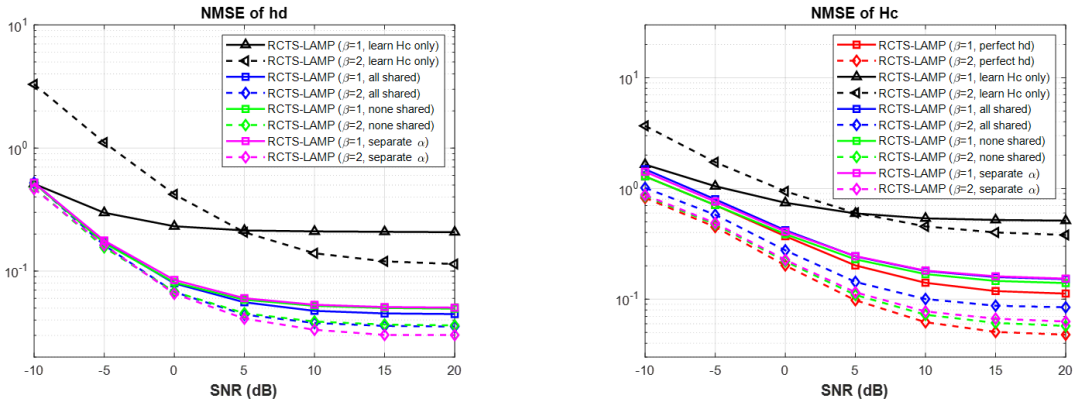
Table II shows that the TS-LAMP network effectively reduces the execution time by 83% at $\beta = 1$ from 24.5 ms to 4.2 ms compared with the LAMP network. Since the TS-LAMP network converges faster at $\beta = 2$ with only $L = 6$ layers, it reduces the execution time by 86% under better estimation performance compared with the LAMP network. On the other hand, the RCTS-LAMP network can further reduce the computational complexity with row compression, which results in the reduction of execution time by 40% at $\beta = 1$ and 47% under $\beta = 2$.

Since the (RC)TS-LAMP-MMV network adopts the learned vector shrinkage function for better estimation performance, the complexity increase in each iteration is $\mathcal{O}\left((Q+K_B)T^2\right)$, which is caused by the calculation of the Onsager correction $\frac{1}{Q}\mathbf{V}_{i-1}\sum_{j=1}^{K_B}\boldsymbol{\eta}'_{vst}\left(\mathbf{r}_{i,j};\ \lambda_i\right)$ in (30). Although the RCTS-LAMP-MMV network with $\beta = 2$ can achieve the best performance, its execution time (2.9ms) is higher than that of the RCTS-LAMP network with $\beta = 2$ (1.8ms). As a result, considering that the RCTS-LAMP network achieves the lowest execution time and almost the same ARSPR performance as the RCTS-LAMP-MMV network at all SNRs, we can conclude that the RCTS-LAMP with $\beta = 2$ achieves the best trade-off between performance and complexity.

D. Joint Estimation Performance of Direct and Cascaded Channels

In this subsection, we consider the general system where the direct and cascaded channels both exist. We evaluate the joint estimation performance with the proposed RCTS-LAMP network under $T = 32$ channel estimation blocks as shown in Fig. 7. To accentuate the effectiveness of the proposed three-stage training procedure, we consider the RCTS-LAMP network that learns cascaded channel estimation by the first stage of the three-stage training procedure. As shown in Fig. 7, under the learning of the cascaded channel only, the RCTS-LAMP network is incapable of delivering a satisfactory performance for both the direct and cascaded channels. This outcome stems from two reasons: firstly, the first LAMP network of RCTS-LAMP is not optimized for direct channel estimation; secondly, the estimation error of the direct channel is not considered in cascaded channel estimation.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3419845



(a) (b)

Fig. 7. Comparison of joint estimation performance of the RCTS-LAMP network for direct and cascaded channels under $T = 32$ channel estimation blocks: (a) NMSE of $\mathbf{h}_d$ versus SNR, (b) NMSE of $\mathbf{H}_c$ versus SNR.

In contrast, through the three-stage training procedure for the RCTS-LAMP network, the joint estimation performance of both the direct and cascaded channels can be substantially improved under the three weight-sharing methods, i.e., all-shared, none-shared, and separate α methods. From Fig. 7(a), we can observe that both the all-shared and separate α methods can achieve the performance of the none-shared method, which means these two weight-sharing methods do not cause performance loss in direct channel estimation. However, in the case of cascaded channel estimation with $\beta = 2$, as illustrated in Fig. 7(b), the all-shared method is unable to attain the performance of none-shared and separate α methods. These two methods nearly achieve the performance of the perfect $\mathbf{h}_d$ scenario. The perfect $\mathbf{h}_d$ scenario refers to the case where the direct channel is blocked or perfectly estimated. Since the perfect $\mathbf{h}_d$ scenario is free from the estimation error of $\mathbf{h}_d$, it establishes an upper bound for cascaded channel estimation performance.

Fig. 8 shows the comparison of the total NMSE, which is the sum of the NMSEs for $\mathbf{h}_d$ and $\mathbf{H}_c$. Since the final objective of the training procedure is the joint NMSE loss $\mathcal{L}_{JO}$, the none-shared method achieves the lowest total NMSE as expected. Note that we can use the same linear transform matrices in the first LAMP network without performance loss since both $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ are measured by the same matrix $\mathbf{A}_1$. Therefore, the separate α and all-shared methods can achieve slightly better direct channel estimation performance by learning how to jointly recover $\mathbf{H}_{row}$ and $\widetilde{\mathbf{h}}_d$ with the same linear transform matrices. On the other hand, the separate α method improves the cascaded channel estimation performance with separate shrinkage parameters. As a result, the separate α method results in the most cost-effective model since it incurs negligible memory overhead compared with the all-shared method while improving the channel estimation performance.

VI. Conclusion

In this paper, we present a novel RCTS-LAMP network for the joint estimation of direct and cascaded channels in IRS-aided mmWave systems. With the two-stage recovery process and the row compression mechanism, the RCTS-LAMP network can significantly reduce the computational complexity

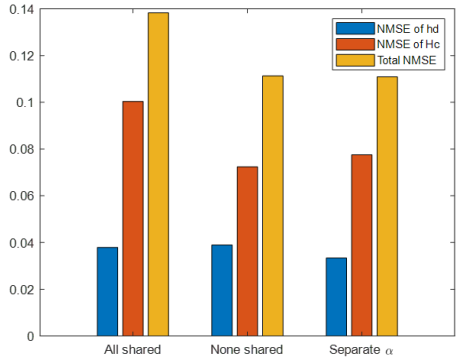


Fig. 8. Joint estimation performance of the RCTS-LAMP network under different weight-sharing methods when $T = 32$ and SNR $= 10$ dB.

compared with the joint CS recovery method. Although the RCTS-LAMP-MMV network solves the recovery problem from the MMV perspective, the RCTS-LAMP network can achieve its performance with a much lower execution time. Moreover, the proposed three-stage training procedure makes the RCTS-LAMP network capable of recovering the direct and cascaded channels. Therefore, the proposed RCTS-LAMP network can achieve a better trade-off between computational complexity and estimation accuracy for cascaded channel estimation, while avoiding extra network for direct channel estimation.

References

[1] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 581–585, May 2020.

[2] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[3] S. Gong *et al.*, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2283–2314, Jun. 2020.

[4] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[5] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14 960–14 973, Dec. 2020.

[6] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[7] A. Bereyhi, S. Asaad, C. Ouyang, R. R. Müller, R. F. Schaefer, and H. V. Poor, "Channel hardening of IRS-aided multi-antenna systems: How should IRSs scale?" *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2321–2335, Aug. 2023.

[8] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 5000–5004.

[9] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[10] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020.

[11] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Process. Lett.*, vol. 28, pp. 513–517, Feb. 2021.

[12] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.

[13] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[14] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020.

[15] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6853–6869, Oct. 2023.

[16] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, Sep. 2021.

[17] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.

[18] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 471–487, Mar. 2021.

[19] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 898–912, Feb. 2022.

[20] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep multi-stage CSI acquisition for reconfigurable intelligent surface aided MIMO systems," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2024–2028, Jun. 2021.

[21] Y. Wang, H. Lu, and H. Sun, "Channel estimation in IRS-enhanced mmWave system with super-resolution network," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2599–2603, Aug. 2021.

[22] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5635–5648, Aug. 2020.

[23] B. Dai, R. Liu, and Z. Yan, "New min-sum decoders based on deep learning for polar codes," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2018, pp. 252–257.

[24] Y. Jiang, S. Kannan, H. Kim, S. Oh, H. Asnani, and P. Viswanath, "DEEPTURBO: Deep turbo decoder," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[25] W.-C. Tsai, C.-W. Chen, and A.-Y. A. Wu, "Compressive channel estimation for IRS-aided millimeter-wave systems via two-stage LAMP network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[26] M. Borgerding and P. Schniter, "Onsager-corrected deep learning for sparse linear inverse problems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 227–231.

[27] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[28] K. Chang, S. I. Kwak, and Y. J. Yoon, "Equivalent circuit modeling of active frequency selective surfaces," in *Proc. IEEE Radio Wireless Symp.*, Jan 2008, pp. 663–666.

[29] W.-C. Tsai, C.-W. Chen, C.-F. Teng, and A.-Y. Wu, "Low-complexity compressive channel estimation for IRS-aided mmwave systems with hypernetwork-assisted LAMP network," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1883–1887, Aug. 2022.

[30] J. Johnston and X. Wang, "Model-based neural networks for massive and sporadic connectivity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 964–969.

[31] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2019, pp. 735–740.

Wen-Chiao Tsai (S'23) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2020, where he is currently pursuing the Ph.D. degree with the Graduate Institute of Electronics Engineering. His research interests are in the areas of machine learning-assisted wireless communication systems design, IRS-assisted wireless communication systems design, and VLSI architecture for DSP.

Chi-Wei Chen (S'23) received his B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2020. He is currently pursuing a Ph.D. degree in the Graduate Institute of Electronics Engineering, National Taiwan University. His research interests are in the areas of IRS-assisted wireless communication systems design, and VLSI architecture for DSP.

An-Yeu (Andy) Wu (M'96-SM'12-F'15) received the B.S. degree from NTU in 1987 and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1992 and 1995, respectively, all in electrical engineering. In August 2000, he joined as a Faculty Member of the Department of Electrical Engineering and the Graduate Institute of Electronics Engineering, NTU, where he is currently a Distinguished Professor. His research interests include low-power/high-performance VLSI architectures and IC designs for DSP/communication/AI applications and adaptive/bio-medical signal processing. From 2020 to 2021, he served as the Editor-in-Chief (EiC) of IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS (JETCAS).